

 04

Teste unicaudal

Transcrição

[0:00] Maravilha, pessoal. Continuando com o teste unicaudal, a gente já conheceu a distribuição T de Student, que é o que a gente vai utilizar para resolver esse probleminha aqui.

[0:08] Já está aqui, a gente leu no vídeo anterior. É aquele problema de que o fabricante alega que determinado produto dele tem, no máximo, 37 gramas de açúcar.

[0:18] Não passa disso. Ou seja, aqui ele já deu a hipótese nula para a gente, a igualdade está ali. E reparem, também está no material de estudo aqui de vocês, que geralmente a hipótese que a gente está testando.

[0:30] Ela fica configurada como a hipótese alternativa. No caso aqui, o que eu estou testando? Se essa alegação desse cara é mentira. Eu acho que tem mais de 37 gramas. É isso que eu vou testar.

[0:44] A tabelinha está aqui, a gente vai continuar teste unicaudal. Aqui os tipos de testes. No caso, aqui H1, que é a hipótese alternativa, que é do tipo maior que, a gente tem um teste unicaudal superior.

[1:00] Onde a área de rejeição está aqui na cauda superior. Rejeição de H0. E a área de aceitação é essa daqui, um menos Alfa. O outro tipo é o H1 do tipo menor que.

[1:13] Onde a área de rejeição está aqui na cauda inferior. Aqui o Alfa, aqui o um menos Alfa, que aqui é a área de aceitação de H0.

[1:22] Esse tipo de teste, eles verificam as variáveis em relação a um piso e a um teto. É importante aqui verificar justamente essas coisas que a gente está vendo no nosso problema.

[1:31] O cara está alegando que é no máximo 37, então o problema está passado daquele valor. Aqui no outro é o contrário. Perfeito? Então vamos lá.

[1:41] Os dados do problema estão aqui, do mesmo jeito que a gente fez no outro. Então, vamos rodar aqui. Isso aqui são as 25 latinhas que o cara foi lá no supermercado selecionar e fez as medições de açúcar.

[1:55] Igual ao outro, eu coloquei essa amostra dentro de um Dataframe. Estão aqui as informações, perfeito. Eu vou calcular a média amostral, tudo bem? 36,25.

[2:06] Também vou calcular o desvio padrão da amostra, porque não me foi dado o desvio padrão da população. Tenho que pegar a amostra e calcular o desvio padrão.

[2:15] Para poder obter a estatística de teste. O procedimento é basicamente o mesmo do que a gente fez antes. Novamente, a média aqui é 37, é o que o cara está alegando lá, é o que a gente está testando.

[2:28] Significância é 5%. Confiança é um menos a significância. O N, 25 latinhas que foram selecionadas. Agora a gente vai utilizar os graus de liberdade porque a gente está utilizando o T de Student.

[2:40] A gente já viu que grau de liberdade é igual a N menos um. Ou seja, 25 menos um, 24. Rodou tudo isso daqui. Maravilha, tudo pronto.

[2:49] As hipóteses a gente praticamente já falou. O que o cara está falando aqui é que o quê? O refrigerante dele tem no máximo 37. Ou seja, pode ter um pouco menos, mas não passa de 37, segundo ele.

[3:01] Ele está afirmando isso. O H1 que a gente está testando aqui é que a média é maior que 37.

[3:08] Vou testar essa minha amostra a um nível de significância de 5% e vou ver se eu consigo resolver esse meu problema, passar por essas hipóteses aqui.

[3:16] Passo dois, a gente fez no programa anterior, onde a gente chegou à conclusão de que o N é maior que 30. Aqui a gente tem a informação no problema que se distribui como uma normal.

[3:27] A gente não conhece o Sigma, então a gente chegou a esse ponto aqui onde a gente tem que usar o T, que é a estatística que vem do T de Student, e o S, que é o desvio padrão da amostra.

[3:37] Está aqui tudo certinho. Vamos lá. Vamos pegar e achar o valor crítico ali, o T Alfa.

[3:44] Alfa por quê? Antes a gente fez o Z Alfa sobre dois porque era bicaudal, a gente tem que dividir o Alfa em duas partes, é simétrica.

[3:51] Agora não, o Alfa está de um lado só, por isso eu vou chamar de T, porque a gente está usando o T de Student, Alfa.

[3:57] Isso aqui já foi importado num vídeo anterior, mas eu vou deixar aqui para deixar bem claro. From scipy.stats import T as t_student.

[4:08] Essa que é a função, a T. Eu não vou usar ela porque eu vou chamar a minha estatística de T e se eu fizer isso vai ter conflito entre a funcionalidade do Python e a minha variável.

[4:18] Então eu dei um apelido para ela, as t_student. Toda vez que eu chamar essa função eu vou ter que chamar por esse nome. Perfeito? É só isso.

[4:26] Aqui eu trouxe já, já deixei aqui pronto parte da tabela que a gente construiu no vídeo anterior, onde tem justamente os graus de liberdade que eu estou interessado.

[4:37] Ou seja, 24 é o grau de liberdade que eu estou interessado. Eu tenho o que? Um teste unicaudal, como a gente já percebeu aqui. Desculpa, eu vim para cima, foi para baixo.

[4:49] Ele vai ser um teste desse tipo aqui. E o que eu tenho que fazer? Vir aqui na minha tabela e ver a linha unicaudal. Qual é o nível de significância? 5%, 0,05, é esse cara aqui.

[5:02] O meu grau de liberdade está aqui, então o meu T vai ser 1,71088. Perfeito? É isso. Na tabela eu já resolvi esse problema aqui. Está aqui, 1,71.

[5:13] Mas tem como a gente fazer isso também utilizando o nosso amigo Python. t_alpha - lembra que não tem mais o Alfa sobre dois - vai ser igual.

[5:23] Aí eu chamo, vou copiar aqui para ficar bem claro, t_student, do mesmo jeito que a gente faz com a normal, ponto PPF. Só que aqui eu passo o que?

[5:34] A confiança - já vou explicar porque - e também graus de liberdade, que a gente já rodou.

[5:43] Por que a confiança? Verifiquem aqui. Essa função, ela vai escolher esse têzinho aqui, mas ela vai me passar o quê?

[5:52] Eu tenho que passar essa probabilidade aqui como a normal. Ela vem desse ponto até o menos infinito, por isso que eu estou passando a 95%. Aqui, já está pronto aqui, confiança 95%.

[6:04] Desse ponto até aqui, eu tenho que passar a probabilidade com o grau de liberdade que eu estou assumindo, para que ele me retorno esse 1,71 aqui.

[6:14] Então, vamos lá. T Alfa, é igual a 1,71. E já está desenhado aqui, já definimos a área de rejeição. A área de rejeição e a área de aceitação, ok?

[6:26] Cálculos da estatística de teste, exatamente igual a gente fez no vídeo anterior. Só que aqui eu vou chamar ele de T. T vai ser igual.

[6:32] Eu vou abrir e fechar aqui, porque tem um numerador e um denominador. Primeiro aqui media_amostra menos a média.

[6:46] Aqui, desvio_amostra, dividido por Numpy.sqrt - função para extrair a raiz quadrada de um número - N. Eu vou ter um T de menos 3,876. Eu arredondei isso aqui, menos 3,88.

[7:11] E aqui eu já posicionei ele. Como a gente já está pegando os macetes, a gente já percebe que ele está na área de aceitação de H0.

[7:20] Ou seja, o nosso problema de teste de hipóteses já está resolvido visualmente. Mas vamos supor que a gente não tenha esse recurso gráfico aqui, a gente quer ter certeza.

[7:28] A gente vem aqui no unicaudal superior, esse é o passo cinco. Unicaudal superior; aqui onde estão as hipóteses, igual a nossa lá em cima. A gente obteve essa estatística aqui.

[7:41] Tirando o Sigma, usando o S. E a gente tem que vir aqui, esse é o crítico aqui para o T. Ou seja, rejeitar o H0 se T, que é a estatística que a gente calculou, esse menos 3,88.

[7:55] For menor ou igual a T Alfa. Desculpa, falei errado. T maior ou igual. A pergunta é: T, você é maior ou igual a T Alfa? Então vamos rodar isso daqui.

[8:09] Ele vai dizer que não, é falso isso, ou seja, eu não posso rejeitar H0. Ou seja, eu já conlcuo que o fabricante não está mentindo.

[8:19] Deixei aqui embaixo: "com um nível de confiança de 95% não podemos rejeitar H0, ou seja, a alegação do fabricante é verdadeira."

[8:27] Lógico, baseado na amostra que eu selecionei e nesse nível de confiança, ou nível de significância. Próximo vídeo eu vou mostrar para você, como a gente fez antes, o P valor.

[8:41] Como calcular. E também uma forma de calcular esse Ttest. A gente está fazendo agora o Ttest, a gente fez o Ztest antes. Calcular o Ttest utilizando ferramentas do Statsmodels.

[8:51] Beleza? Até lá.